Russian Entomol. J. 15(2): 263–268

3I, a new program for creating Internet-accessible interactive keys and taxonomic databases and its application for taxonomy of Cicadina (Homoptera)

3I — новая программа для создания интерактивных определительных таблиц и таксономических баз данных для Интернета и её применение в систематике цикадовых (Homoptera: Cicadina)

Dmitry A. Dmitriev
Д.А. Дмитриев

Illinois Natural History Survey, 1816 South Oak Street, Champaign IL 61820, USA. E-mail: proceps@mail.ru
Служба естественной истории Иллинойса, 1816 ул. Саус Оук, Чэмпейн, Иллинойс 61820, USA

KEY WORDS: Web, on-line, key, identification, leafhoppers, Hemiptera.

КЛЮЧЕВЫЕ СЛОВА: Сеть, интерактивный, определительная таблица, определение, цикадки, Hemiptera.

ABSTRACT. The paper describes 3I, a new program for creating Internet-accessible interactive keys and taxonomic databases [http://ctap.inhs.uiuc.edu/dmitriev/]. The character matrix for the Key is stored in a Microsoft Access 2000 database. The action of the Key is based on the capability of Active Server Pages to query the database and retrieve the names of taxa matching the states specified by the user. If multiple taxa are retrieved, the system offers a reduced list of characters relevant to further identification, which enables the user to proceed step by step starting with any character or character combination. The program can work with an unlimited number of taxa as well as with an unlimited number of characters and character states, and is capable of supporting multiple languages and images. The 3I can also be used for supporting on-line searchable taxonomic databases.

РЕЗЮМЕ: В статье приводится описание работы с новой программой 3I, предназначенной для создания интерактивных определительных таблиц и таксономических баз данных для Интернета [http://ctap.inhs.uiuc.edu/dmitriev/]. Матрица признаков для определительной таблицы хранится в базе данных Microsoft Access 2000. Работа определительной таблицы основана на возможности создавать запросы к базе данных через Активные Серверные Страницы (ASP) и возвращать список названий таксонов, соответствующих состояниям признаков, выбранным пользователем. Если в списке остаётся больше одного таксона, то программа переоценивает список признаков и оставляет только те из них, которые пригодны для дальнейшего определения оставшихся таксонов. Это позволяет пользователю шаг за шагом продолжать определение, начиная с любого признака или комбинации признаков. Программа рассчитана на работу с неограниченным количеством таксонов, признаков и их состояний, в ней заложена возможность работы с иллюстрациями и поддержка интерфейса на нескольких языках. 3I также может быть использована для создания таксономических баз данных с возможностью расширенного поиска в режиме онлайн.

Introduction

Recently, a number of keys for identification of various groups of living organisms have appeared on the Internet [see, for example, collections of links on the sites of M.B. Dianov and A.L. Lobanov (???? onwards); H. Berger (1998 onwards)]. It is much quicker and easier to publish on the Internet than in a journal, moreover, such resources can be easily accessed by various categories of users. Most available keys are not interactive, but have the same dichotomous structure as traditionally used in printed keys.

Any identification key is based on a matrix, in which states of all characters are recorded for every taxon (in some cases only in the mind of the author). Such a matrix can be directly used as a tabular key. In order to identify a specimen, its features should be compared with each line of the matrix. A few Internet keys, based on this principle, do the comparison automatically [e.g., Hill, 1998 onwards; McAleece & Doolan, 2003 onwards; Wang, 2000 onwards]. Only one step of identification is allowed: the user specifies all or part of the characters, and the program retrieves one or several matching taxa. This type of key might be referred to as semi-interactive.

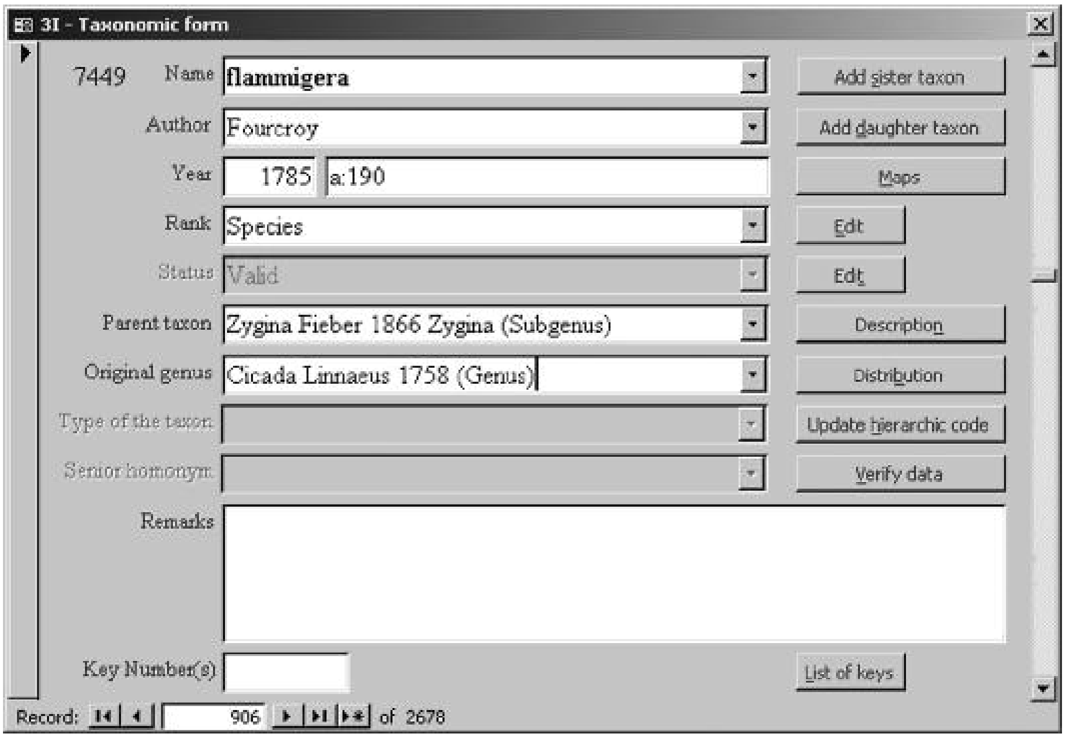


Fig. 1. 3I database. Taxonomic form.
Рис. 1. База данных 3I. Таксономическая форма.

Tabular keys have certain limits for the number of characters and, in printed versions, for the number of taxa, beyond which using the key becomes cumbersome. The traditional way to deal with a large data matrix is to convert it into a dichotomous key. While this simplifies identification, much of the data is lost. Moreover, identification must always start from the same point. If, for example, the key starts with characters of a front leg, and the specimen has both front legs missing, identification becomes difficult.

True interactive keys overcome the imperfections of dichotomous keys. Like tabular keys, they use all the data in a character matrix. A user can start with any character and ask the computer to retrieve the list of taxa in which this character is present, while at the same time reducing the list of characters to those relevant for further identification. The avoidance of fixed sequences of characters like those in dichotomous keys allows the user to start identification with the features he or she can see more easily. If the data matrix is sufficiently redundant, taxa can be identified even if some structures are missing or poorly developed [for more detailed description of interactive key advantages see Dallwitz, 2000; Dallwitz, Paine & Zurcher, 2003a]. There are several programs which allow users with no or only minimal programming skills to create interactive keys [reviewed by Dallwitz, 2002]. To build such a key, one only needs a list of taxa, a list of characters, a data matrix (like a standard matrix for phylogenetic analysis), and perhaps some illustrations. The programs generating interactive keys can be divided into two major categories. The first one includes programs that make keys which can be distributed on CDs, other media, or via the Internet and can be run on a local computer, for example, Delta IntKey [Dallwitz, Paine & Zurcher, 1993, 1993 onwards], LucID [Young, Thiele & Yeats, 1997; Centre for Biological Information Technology…, ???? onwards], PICKEY [Dianov, Lobanov, 1997, ???? onwards]. To work with them, special software must be installed on the user's computer. The programs of the second group generate Internet-accessible keys, for example, 20q [Pickering, Polistes Corporation, ???? onwards], SLIKS [Guala, 2004 onwards]. Nothing has to be installed on user's computer to work with them; they are much less dependant on the computer platform. On the other hand, in most cases, keys of the second category can not be started on a local computer off-line [for more details about advantages and disadvantages of Internet-accessible keys see Dallwitz, Paine & Zurcher, 2003b; Lobanov, 2003 onwards].

The present paper describes 3I[1], a new program for creating Internet-accessible interactive keys and taxonomic databases [Dmitriev, 2003 onwards].

Program description

The 3I key is based on a character state matrix stored in a MS Access 2000 database (Figs 1–2). The action of the program is based on the capability of ASP-pages (Active Server Pages) to query the database and re-

[1] Internet-accessible Interactive Identification

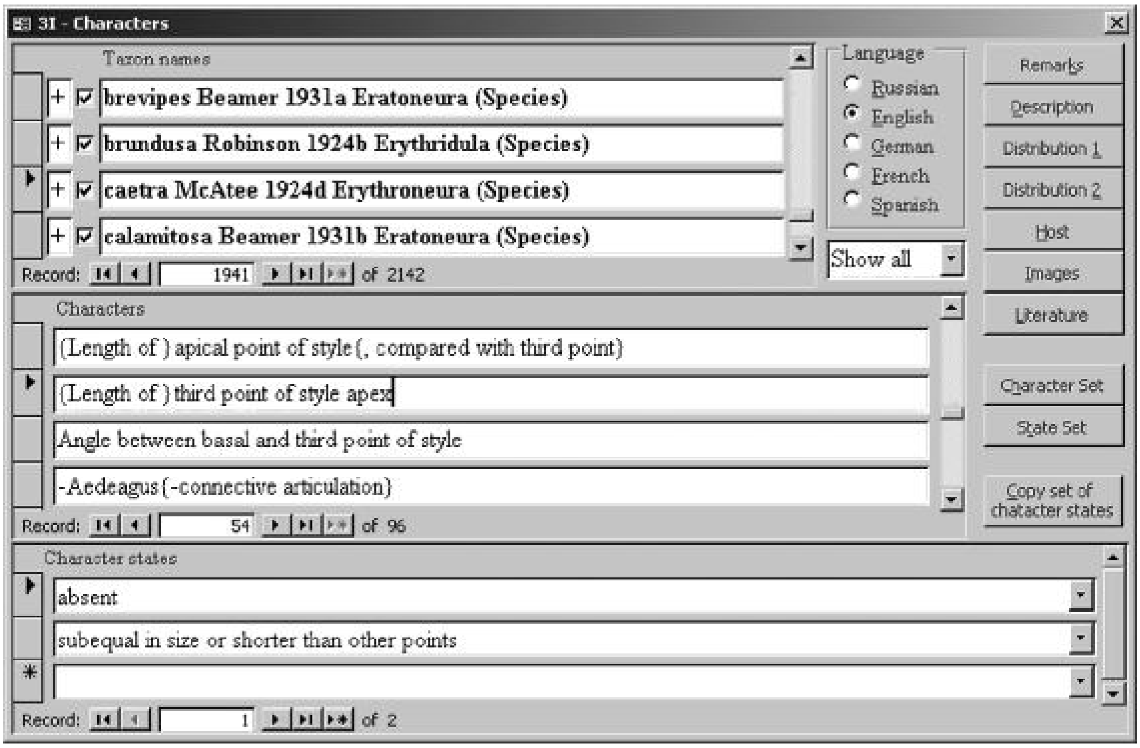


Fig. 2. 3I database. Character form.
Рис. 2. База данных 3I. Форма с признаками.

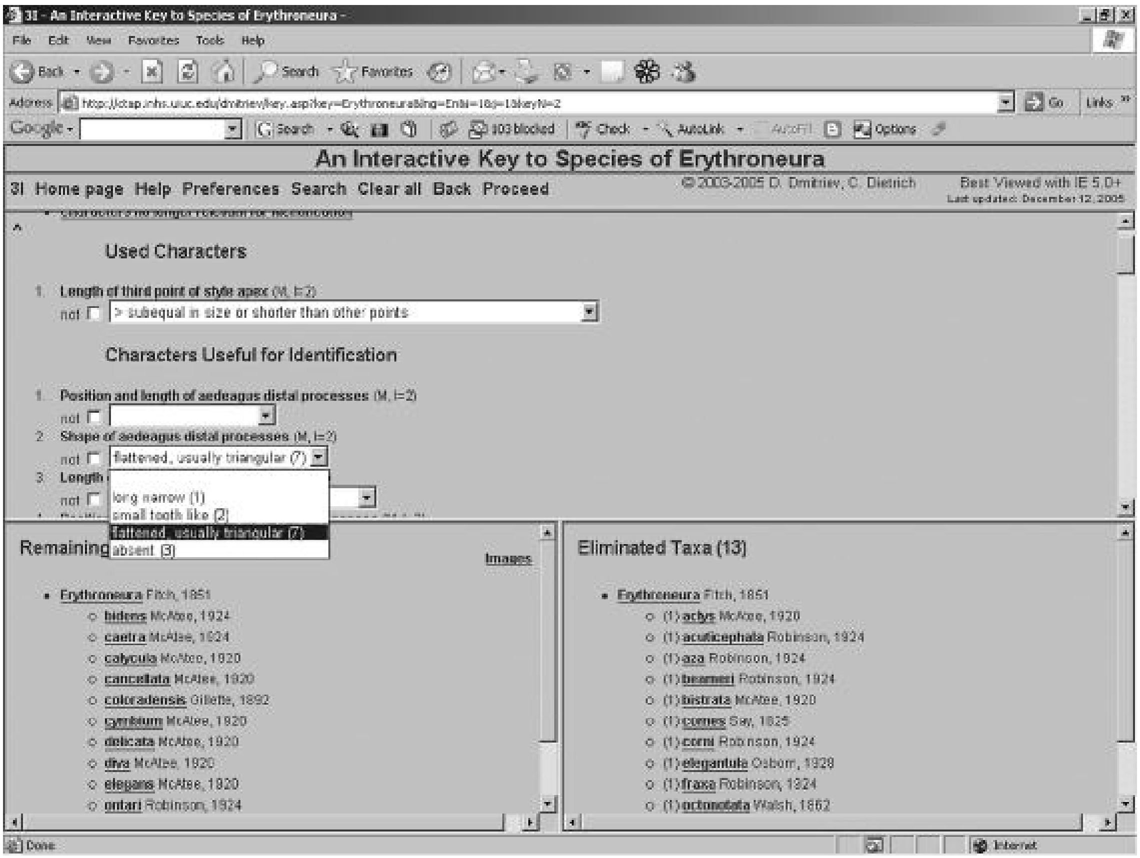


Fig. 3. 3I interface (screen resolution 1024x768).
Рис. 3. Интерфейс определительной таблицы 3I (разрешение экрана 1024x768).

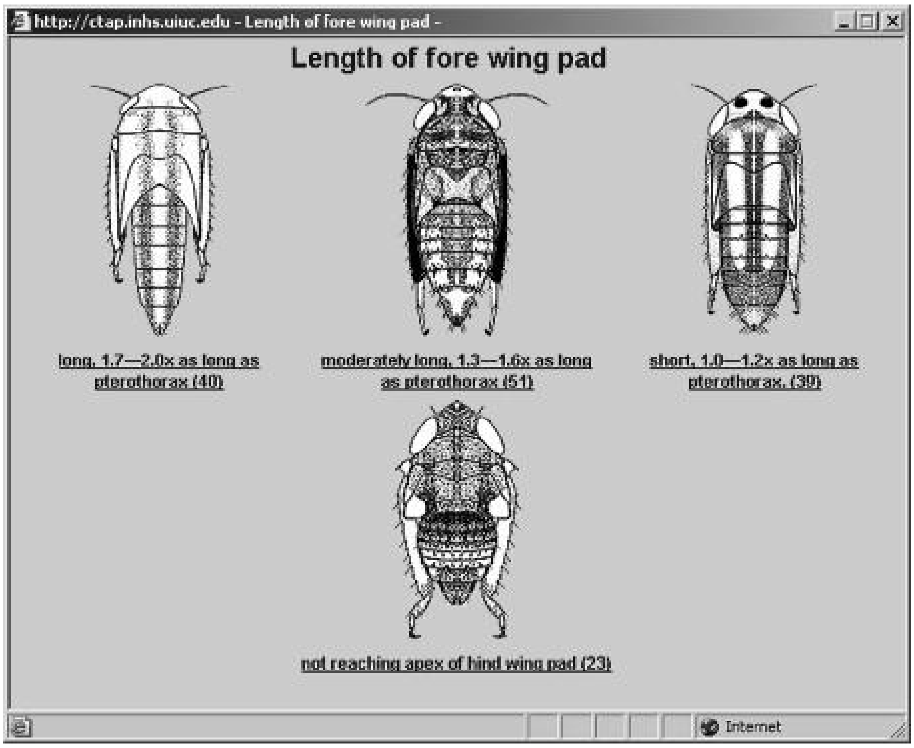


Fig. 4. 3I. Example of explanatory images for character states.
Рис. 4. 3I. Пример с поясняющими рисунками для состояний признака.

trieve the names of taxa matching the character states specified by the user. If a character is not scored for a taxon, this taxon will remain in the list regardless of which character state is chosen.

The main work field (Fig. 3) presents the list of characters. Each character has a drop-down box with a set of states. The characters can be sorted in three ways. 1. *By rank*. The author of the Key specifies which characters are the best for identification (e.g., those easiest to see and/or least likely to be misinterpreted). The rank is displayed in parentheses after the character name. 2. *By separating power*. The author-assigned ranks are disregarded and the program calculates which characters best divide the taxa into groups of equal size (recalculated after each step of identification). 3. *By morphology*. Characters are listed according to body region, e.g., head, thorax, etc., and sorted by rank within each region.

Identification begins by choosing the state of any character from the drop-down box. Some of the characters also have a hyperlink to explanatory images (Fig. 4). After the states of one or several characters have been chosen, the **Proceed** button should be pressed in order to renew the list of taxa matching the search criteria, the list of eliminated taxa, and to limit the list of characters and their states to those useful for further identification. Character states not relevant for further identification are not removed, but marked with minus symbol. The number after each state shows how many of the remaining taxa have that state. If a character has multiple states and there is uncertainty which one to chose, the NOT option can be helpful; the taxa not having the selected state will remain in the list. For continuous or numeric characters, no states are listed; the user should enter the exact numeric value into the field (the range of valid values is shown in parentheses after the character name). The names of taxa are hyperlinked to verbal taxon descriptions generated automatically from the character data. The button **Clear all** is used to return the character set to the original list and to start identification from the beginning.

The **Preferences** button allows one to modify the interface and the set of characters displayed. For example, if a specimen to be identified is a female, the user can choose to display only characters applicable to females by deselecting the check box next to “Males”. The applicability of each character based on stage/sex is shown in parentheses together with the importance index: M — male characters, F— female characters, N — nymphal characters, L — larval characters, P — pupal characters, E — egg characters, and D — damage characters. If the Internet connection is slow, the user can slightly increase the download speed by choosing not to display eliminated characters and/or taxa. The preferences also include options for sorting the character list described above, and for changing the error tolerance via a drop-down box. The latter option is useful if, after several steps of identification, no taxon remains that

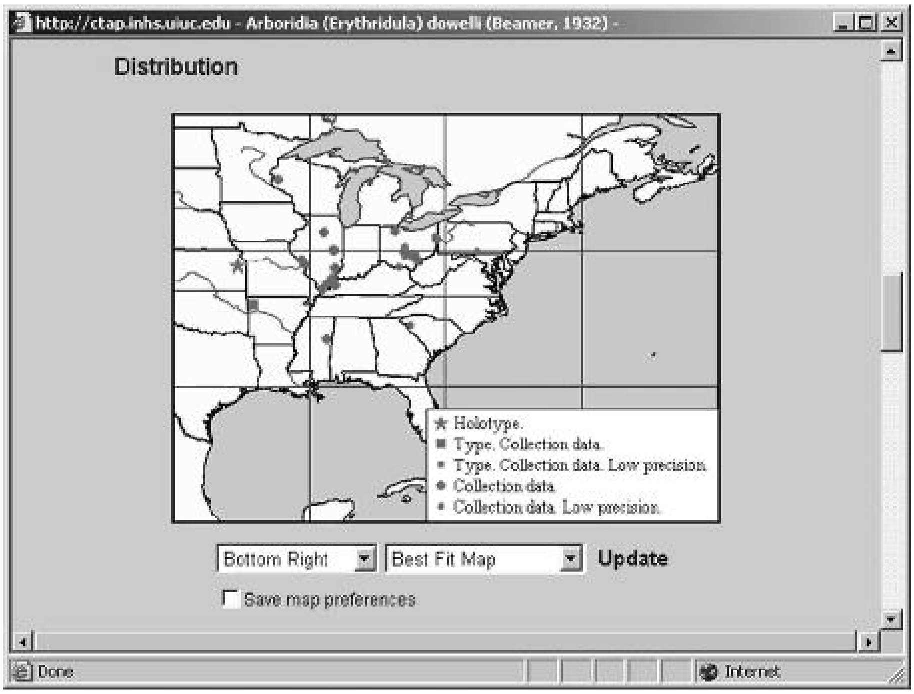


Fig. 5. 3I. Example of interactive map.
Рис. 5. 3I. Пример с интерактивной картой.

matches the criteria of search. In this case, the user can either go one or more steps back and try other states or characters, or increase the error tolerance to a value greater than the default of "0". The error tolerance is the number of mismatching characters allowed during identification. When this option is used, after each step of identification the number of mismatching characters is displayed in parentheses before the taxon names. Increasing the error tolerance relaxes the criteria of search and increase the list of matching taxa. Taxa are then eliminated when the mismatch exceeds the specified error tolerance. Other optional features are Language switch and Rank limit. The latter allows the user to limit identification to higher taxa (for example, if the key is for identification of species, the list of taxa can be limited to genera or tribes).

In addition to supporting interactive keys, the program can be used to store a variety of taxonomy-related information: synonymic and common names, images, distribution (including generation of interactive maps, see Fig. 5), hosts, literature etc. This information is linked to an interactive key or can be retrieved through the Search window. With these features, the 3I comprises a nearly complete set of tools for producing taxonomic revisions.

The 3I was designed to work with any browser supporting frames and JavaScript. But Internet Explorer 5 or later version is recommended for better performance. Other browsers may cause undesirable changes in the interface. So far, the program was tested and works correctly with Internet Explorer 5.0–6.0, Netscape 4.08–7.02, Opera 5.0–7.2, Safari, Mozilla, Firefox, and MSN browsers on both Mac and PC computers. Some users, however, have noted unstable performance with earlier versions of Netscape 4.

3I application for taxonomy of Cicadina

The 3I program has already been used to create several interactive keys and taxonomic databases, which are listed below:

Family Cicadellidae (by D. Dmitriev).

- Interactive key to leafhopper tribes and subfamilies based on adult and nymph morphology.
- List of literature on leafhopper nymphs (115 names)

Tribe Erythroneurini (by D. Dmitriev & C. Dietrich).

- Interactive key to genera of the tribe Erythroneurini.
- Interactive key to species of the subgenus *Erythroneura*.
- Interactive key to species of the subgenus *Eratoneura*.
- Interactive key to species of the subgenus *Erasmoneura*.
- Interactive key to species of the subgenus *Erythridula*.
- Interactive key to species of the genus *Hymetta*.
- Taxonomic database on the tribe Erythroneurini (2678 names, including 2105 valid names, distribution and host plant information for most of the species, including geo-references for over 80 000 specimens of the genus *Erythroneura* from collections throughout the U.S. and Canada, and more than 6850 images, mostly for the genus *Erythroneura*, list of literature — 313 names).

Tribe Proconiini (by D. Takiya & D. Dmitriev).
- Interactive key to the genera of the tribe Proconiini.
- Taxonomic database on the tribe Proconiini (684 names, including 488 valid names, distribution and host plant information for most of the species, 200 images).

World Cicadellinae and Typhlocybinae (by S. McKamey).
- Taxonomic database (9245 taxa, including 7930 valid names).

New World Membracidae (by C. Dietrich).
- Interactive key to New World membracid tribes.

North American Fulgoroidea (by S. Wilson & L. O'Brien).
- Taxonomic database on North American Fulgoroidea (1194 names, distribution for all species within U.S.)

ACKNOWLEDGEMENTS. I am thankful to Chris Dietrich, Roman Rakitov, Daniela Takiya (all from Illinois Natural History Survey, USA), Mike Dallwitz (CSIRO Entomology, Australia), and Joanna McCaffrey (Chicago Field Museum, USA) for critical review and discussion of 3I's structure. The work was supported in part by NSF grant DEB-0315373.

References

Berger H. 1998. onwards. Helmut Berger — Home page. http://www.protozoology.com/

Centre for Biological Information Technology of The University of Queensland. Lucidcentral Home. http://www.lucidcentral.com/

Dallwitz M.J. 2000. Desirable attributes for interactive identification programs. http://delta-intkey.com/www/idcriteria.htm

Dallwitz M.J. 2002. Programs for Interactive Identification and Information Retrieval. http://biodiversity.uno.edu/delta/www/idprogs.htm

Dallwitz M.J., Paine T.A. & Zurcher E.J. 1993. User's guide to the DELTA System: a general system for processing taxonomic descriptions. 4th ed. Canberra, Austria: CSIRO. 133 p.

Dallwitz M.J., Paine T.A. & Zurcher E.J. 1993 onwards. DELTA. Http://delta-intkey.com/

Dallwitz M.J., Paine T.A. & Zurcher E.J. 2003a. Principles of interactive keys. http://delta-intkey.com/www/interactivekeys.htm

Dallwitz M.J., Paine T.A. & Zurcher E.J. 2003b. Interactive identification using the Internet. http://delta-intkey.com/www/netid.htm

Dianov M.B. & Lobanov A.L. 1997. PICKEY — a program for identification of organisms with interactive use of images. Tr. Zool. Inst. Vol.269. P.35–39. [In Russian].

Dianov M.B. & Lobanov A.L. PICKEY. http://www.zin.ru/projects/pickey/

Dmitriev D.A. 2003 onwards. Dmitry A. Dmitriev — Home page. http://ctap.inhs.uiuc.edu/dmitriev/

Guala G.F. 2004 onwards. SLIKS (Stinger's Lightweight Interactive Key Software). http://www.stingersplace.com/SLIKS/

Hill K. 1998 onwards. The cycad pages. http://plantnet.rbgsyd.gov.au/PlantNet/cycad/

Lobanov A.L. 2003 onwards. Keys for beetles and biological diagnostics. http://www.zin.ru/animalia/coleoptera/rus/syst8.htm

McAleece N. & Doolan S. 2003 onwards. Taxonomic information across the Internet. http://www.nhm.ac.uk/zoology/taxinf/

Pickering J. & Polistes Corporation. ID nature guides Discover life. http://pick4.pick.uga.edu/mp/20q

Wang X.P. 2000 onwards. Coelotinae. http://coelotinae.inhs.uiuc.edu/coelotinae/

Young T., Thiele K. & Yeats D. 1997. LucID player instruction manual. Brisbane: Coop. Res. Centre Tropical Pest Management, Univ. Queensland. 83 p.